UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 13-22
Statistics and Econometrics Series 20
June 2013

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A multivariate extension of a vector of Poisson-Dirichlet processes

W. Zhu [1] and F. Leisen [2],

Abstract

Recently, Leisen and Lijoi (2011) introduced a bivariate vector of random probability measures with Poisson-Dirichlet marginals where the dependence is induced through a Lévy's Copula. In this paper the same approach is used for generalizing such a vector to the multivariate setting. Some non-trivial results are proved in the multidimensional case, in particular, the Laplace transform and the Exchangeable Partition Probability function (EPPF). Finally, some numerical illustrations of the EPPF are provided.

Keywords: Bayesian inference, Dirichlet process, vectors of Poisson-Dirichlet processes, Multivariate Lévy measure, Partial exchangeability, Partition probability function

[1] Weixuan Zhu, Departamento de Estadistica, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain. Email: zhuweixuan@gmail.com .

[2] Fabrizio Leisen, Departamento de Estadistica, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain. Email: fabrizio.leisen@gmail.com . The research of Fabrizio Leisen has been partially supported by the Spanish Ministry of Science and Innovation through grant ECO2011-25706.

1. Introduction

Random probability vectors are of great interest, especially in view of their application to statistical inference. Recently, a growing literature in Bayesian non-parametrics proposed new priors for modeling situations where data may be divided into different groups. In this case, one would like to consider different densities for different groups instead of a single common density for all the data. For this reason, models driven by vectors of dependent random probability measures could be used as an alternative to the classical univariate models. After the seminal papers of MacEachern (1999), the problem of modelling a finite number of dependent densities, allowing information pooling across groups, has become an active area of research in Bayesian non-parametrics. Focusing on the stick breaking representation of the Dirichlet Process (DP), De Iorio et al. (2004) propose an ANOVA-type dependence for the law of the atoms, and later, Griffin and Steel (2006), defined a class of DP with both dependent atoms and weights. Many other alternative constructions have been proposed in this framework, see, for instance Hatjispyrosa et al. (2011), Kolossiatis et al. (2013) and the references therein. All these approaches are based on the Sethuraman representation of the DP. In Leisen and Lijoi (2011) an alternative way for constructing vectors of random probability measures is proposed. They introduce a 2-dimensional infinite divisible vector where the dependence is regulated by a Levy's Copula. The same approach is used in Leisen et al. (2013).

In this paper, we extend to the multidimensional case some results presented in Leisen and Lijoi (2011), in particular, the Laplace Transform and the exchangeable partition probability function. In addition, a MCMC algorithm is proposed for evaluating the Exchangeable Partition Probability Function (EPPF). When the dimension is greater than 2, the derivation of the Laplace transform is non-trivial and the result that we prove, it is interesting compared with the Laplace transform of the n-dimensional vector of Dirichlet Processes studied in Leisen et al. (2013). In addition, an MCMC algorithm for evaluating the EPPF is proposed and its performance is tested in the bidimensional case. As we will show, the application of the algorithm to some specific examples, allows some considerations about the clustering behaviour of such a vector. The paper is organized as follows. In Section 2 some preliminaries are given and the vector of Poisson-Dirichlet processes is introduced. Section 3 is devoted to the derivation of the multivariate Laplace exponent and in Section 4 an expression of the EPPF is provided for the multivariate setting. Finally, Section 5 is devoted to the definition of the MCMC algorithm, the testing of its perfomance and the application to some examples.

2. Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathbb{X}, \mathcal{X})$ a measure space, with $\mathbb{X}$ Polish and $\mathcal{X}$ the Borel σ–algebra of subsets of $\mathbb{X}$. Suppose $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are completely random measures[1] (CRMs) on $(\mathbb{X}, \mathcal{X})$ with respective marginal Lévy measures

$$\bar{\nu}_i(\mathrm{d}x, \mathrm{d}y) = \alpha(\mathrm{d}x)\, \nu_i(\mathrm{d}y) \qquad i = 1, \ldots, d$$

The probability measure α on $\mathbb{X}$ is non–atomic and ν_i is a measure on $\mathbb{R}^+$ such that $\int_{\mathbb{R}^+} \min(y, 1)\, \nu_i(dy) < \infty$. We further suppose that $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are stable CRMs, *i.e.*

$$\nu_i(dy) = \frac{\sigma}{\Gamma(1-\sigma)} y^{-1-\sigma}\, \mathrm{d}y \qquad 0 < \sigma < 1 \text{ and } i = 1, \ldots, d. \tag{1}$$

Moreover, $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are dependent and the random vector $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ has independent increments, in the sense that given A and B in $\mathcal{X}$, with $A \cap B = \varnothing$, then $(\tilde{\mu}_1(A), \ldots, \tilde{\mu}_d(A))$ and $(\tilde{\mu}_1(B), \ldots, \tilde{\mu}_d(B))$ are independent. This implies that for any set of measurable functions $\boldsymbol{f} = (f_1, \ldots, f_d)$ such that $f_i : \mathbb{X} \to \mathbb{R}$, $i = 1, \ldots, d$ and $\int |f_i|^\sigma \, d\alpha < \infty$, one has

$$\mathbb{E}\left[e^{-\tilde{\mu}_1(f_1)-\cdots-\tilde{\mu}_d(f_d)}\right] = \exp\left\{-\psi^*_{\rho,d}(\boldsymbol{f})\right\}. \tag{2}$$

where

$$\psi^*_{\rho,d}(\boldsymbol{f}) = \int_{\mathbb{X}} \int_{(0,\infty)^d} \left[1 - \mathrm{e}^{-y_1 f_1(x)-\cdots-y_d f_d(x)}\right] \rho_d(\mathrm{d}y_1, \ldots, \mathrm{d}y_d)\, \alpha(\mathrm{d}x) \tag{3}$$

The representation (2) entails that the jump heights of $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ are independent from the locations where the jumps occur. Moreover, these jump locations are common to all the CRMs and are governed by α.

An important issue is the definition of the measure ρ_d in (2): we will

[1]For background information on CRMs one can refer to Kingman (1993)

determine it in such a way that it satisfies the condition

$$\int_{(0,\infty)^{d-1}} \rho_d(\mathrm{d}x_1, \dots, dx_{j-1}, A, dx_{j+1}, \dots, dx_d) = \int_A \frac{\sigma}{\Gamma(1-\sigma)} y^{-1-\sigma}\,\mathrm{d}y \quad (4)$$

for any $j = 1, \dots, d$ and $A \in \mathcal{B}(\mathbb{R}^+)$. In other words, the marginal Lévy intensities coincide with ν_i in (1). This is possible through a Lévy Copula, see Cont and Tankov (2004). Indeed, set $U_i(y) := \int_y^\infty \nu_i(s)\,\mathrm{d}s$ as the i–th marginal tail integral associated to ν_i and

$$U(y_1, \dots, y_d) = \int_{y_1}^\infty \cdots \int_{y_d}^\infty \rho_d(s_1, \dots, s_d)\,\mathrm{d}s_1 \cdots \mathrm{d}s_d$$

is the corresponding tail integral. According to Cont and Tankov (2004) there exists a unique Lévy copula C such that

$$U(y_1, \dots, y_d) = C(U_1(y_1), \dots, U_d(y_d)).$$

Furthermore, if both the copula C and the marginal tail integrals are sufficiently smooth

$$\rho_d(y_1, \dots, y_d) = \frac{\partial^d C(s_1, \dots, s_d)}{\partial s_1 \cdots \partial s_d}\bigg|_{s_1=U_1(y_1),\dots s_d=U_d(y_d)} \nu_1(y_1) \dots \nu_d(y_d).$$

A choice for C could be the Lévy-Clayton Copula

$$C_\theta(s_1, \dots, s_d) = (s_1^{-\theta} + \cdots + s_d^{-\theta})^{-\frac{1}{\theta}} \qquad \theta > 0. \quad (5)$$

where the parameter θ regulates the degree of dependence. Under the particular choice of stable marginals and $\theta = \frac{1}{\sigma}$, then

$$\rho_d(y_1, \dots, y_d) = \frac{(\sigma)_d}{\Gamma(1-\sigma)} |\mathbf{y}|^{-\sigma-d} \quad (6)$$

where $|\mathbf{y}| = y_1 + \cdots + y_d$ and $(\sigma)_d = \sigma(\sigma+1)\cdots(\sigma+d-1)$ is the ascending factorial. If $d = 2$ then we recover the bivariate Lévy intensity used in Leisen and Lijoi (2011). Let $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ be the vector of random probability measures defined in (2) with ρ_d as in (6). Suppose $\mathbb{P}_{i,\sigma}$ is the probability distribution of $\tilde{\mu}_i$, for $i = 1, \dots, d$.

Hence $\mathbb{P}_{i,\sigma}$ is supported by the space of all boundedly finite measures $\boldsymbol{M}_{\mathbb{X}}$

on $\mathbb{X}$ endowed with the Borel σ–algebra $\mathcal{M}_{\mathbb{X}}$ with respect to the $w^\sharp$–topology ("weak-hash" topology[2]). Introduce, now, another probability distribution $\mathbb{P}_{i,\sigma,\theta}$ on $(\boldsymbol{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ such that $\mathbb{P}_{i,\sigma,\theta} \ll \mathbb{P}_{i,\sigma}$ and

$$\frac{\mathrm{d}\mathbb{P}_{i,\sigma,\theta}}{\mathrm{d}\mathbb{P}_{i,\sigma}}(\mu) = \frac{\Gamma(\theta)}{(K)^{\frac{1}{d}}}[\mu(\mathbb{X})]^{-\theta} \tag{7}$$

where

$$K = \int_{(0,\infty)^d} \left(\prod_{i=1}^{d} \lambda_i\right)^{\theta-1} e^{-\psi_{\rho,d}(\boldsymbol{\lambda})} d\boldsymbol{\lambda}$$

and $\psi_{\rho,d}(\boldsymbol{\lambda})$ is the Laplace exponent defined in formula (9). We denote with $\tilde{\mu}_{i,\sigma,\theta}$ a random element defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(\boldsymbol{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ whose probability distribution coincides with $\mathbb{P}_{i,\sigma,\theta}$. The random probability measure

$$\tilde{p}_i = \tilde{\mu}_{i,\sigma,\theta}/\tilde{\mu}_{i,\sigma,\theta}(\mathbb{X})$$

is a Poisson-Dirichlet process with parameter (σ, θ), see, e.g. Pitman and Yor (1997) and Pitman (2006), and The vector

$$(\tilde{p}_1, \ldots, \tilde{p}_d) \tag{8}$$

is a dependent vector of Poisson–Dirichlet random probability measures on $(\mathbb{X}, \mathcal{X})$. Note that, when $d = 2$, (8) coincides with the vector introduced in Leisen and Lijoi (2011).

Remark. Note that the change of measure in (7) differs from the one given in Leisen and Lijoi (2011). The latter contains a typo that, however, slightly affects only the normalizing constant of the EPPF. Indeed, such a constant must coincide with $\frac{\Gamma(\theta)}{(K)^{\frac{1}{d}}}$.

3. The Laplace Exponent

In this section, it is provided the Laplace exponent of the vector $(\tilde{\mu}_1, \ldots, \tilde{\mu}_n)$ defined through the Lévy intensity in (6). This is an important tool for deter-

[2]Recall that a sequence of measures $(m_i)_{i\geq 1}$ in $\boldsymbol{M}_{\mathbb{X}}$ converges, in the $w^\sharp$–topology, to a measure m in $\boldsymbol{M}_{\mathbb{X}}$ if and only if $m_i(A) \to m(A)$ for any bounded set $A \in \mathcal{X}$ such that $m(\partial A) = 0$. See Daley and Vere-Jones (2003) for further details.

mining the Exchangeable Partition Probability Function in the next section.

Before getting started, it is worth noting that, since $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ has independent increments, its distribution is characterized by a choice of $f_1, \dots, f_d$ in (2) such that $f_i = \lambda_i \, 1_A$ for any set A in $\mathcal{X}$, $\lambda_i \in \mathbb{R}^+$ and $i = 1, \dots, d$. In this case

$$\psi^*_{\rho,d}(\boldsymbol{f}) = \alpha(A)\, \psi_{\rho,d}(\boldsymbol{\lambda})$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ and

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} \left[1 - \mathrm{e}^{-\langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle}\right] \frac{(\sigma)_d}{\Gamma(1-\sigma)} |\boldsymbol{y}|^{-\sigma-d} \mathrm{d}\boldsymbol{y} \tag{9}$$

where $\boldsymbol{y} = (y_1, \dots, y_d)$ and $\langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle = \sum_{i=1}^d \lambda_i y_i$.

In Leisen and Lijoi (2011), the authors provide the expression of $\psi_{\rho,d}$ in the bidimensional case, i.e.

$$\psi_{\rho,2}(\lambda_1, \lambda_2) = \begin{cases} [\lambda_1^{\sigma+1} - \lambda_2^{\sigma+1}]/(\lambda_1 - \lambda_2) & \lambda_1 \neq \lambda_2 \\ (\sigma+1)\lambda_1^{\sigma} & \lambda_1 = \lambda_2 \end{cases} \tag{10}$$

In this case, the computation of $\psi_{\rho,2}$ is quite straightforward but it's not trivial in the multidimensional setting. Our goal is to prove the following

Proposition 1. *Let $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$ be such that it consists of $l \leq d$ distinct values denoted as $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_l)$ with respective multiplicities $\boldsymbol{n} = (n_1, \dots, n_l)$. Then*

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \psi_{\rho,d}(\tilde{\boldsymbol{\lambda}}, \boldsymbol{n}) = \left(\prod_{i=1}^{l} \frac{1}{\Gamma(n_i)} \frac{\partial^{n_i - 1}}{\partial^{n_i - 1} \tilde{\lambda}_i} \right) \left(\phi_l(\tilde{\boldsymbol{\lambda}}) \prod_{i=1}^{l} \tilde{\lambda}_i^{n_i - 1} \right), \tag{11}$$

where

$$\phi_l^{\sigma}(\boldsymbol{x}) = \begin{cases} \sum_{i=1}^{l} \frac{x_i^{\sigma+l-1}}{\prod_{j=1, j\neq i}^{l} (x_i - x_j)} 1_{(x_1 \neq \dots \neq x_l)} & \text{if } l > 1 \\ \lambda_1^{\sigma} & \text{if } l = 1 \end{cases} \tag{12}$$

This result is interesting when compared with the Laplace Exponent of the vector of Dirichlet Processes introduced in Leisen et al. (2013). They differ only in the function ϕ_l^{σ}. Clearly, this is due to the different nature of the marginals but it is surprising if considering that they arise from different Lévy Copulas. Our conjecture is that this happens when the Lévy Copula could be seen as a function of the tail integrals of the marginal Lévy intensities.

The development of this idea (in its generality) is out of the scope of this paper and will be matter of future research.

Before proving the statement of the proposition we need some preliminaries. For every $a > 0$ and $d > 1$ such that $d \in \mathbb{N}$, define the integral

$$\Phi_d^a(\boldsymbol{\lambda}) = \int_{\triangle_{d-1}} (a+1)_{d-1} \left[\lambda_1 z_1 + \cdots + \lambda_{d-1} z_{d-1} + \lambda_d (1 - z_1 - \cdots - z_{d-1})\right]^a d\boldsymbol{z}$$

where $\boldsymbol{z} = (z_1, \ldots, z_{d-1})$ and $\triangle_d = \{\mathbf{z} \in (0,1)^d : z_1 + \cdots + z_d < 1\}$. The Laplace Exponent is strictly related to this integral, indeed a simple change of variable, namely $z_i = y_i / s$ for $i = 1, ..., d-1$ and $s = |\mathbf{y}|$ in equation (9), yields

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \Phi_d^\sigma(\boldsymbol{\lambda})$$

i.e. the Laplace Exponent coincides with the integral (13) when $a = \sigma$. This means that, for proving the statement of Proposition 1, we have to prove that

$$\Phi_d^\sigma(\boldsymbol{\lambda}) = \Phi_d^\sigma(\tilde{\boldsymbol{\lambda}}, \boldsymbol{n}) = \left(\prod_{i=1}^{l} \frac{1}{\Gamma(n_i)} \frac{\partial^{n_i - 1}}{\partial^{n_i - 1} \tilde{\lambda}_i}\right) \left(\phi_l(\tilde{\boldsymbol{\lambda}}) \prod_{i=1}^{l} \tilde{\lambda}_i^{n_i - 1}\right), \qquad (13)$$

Proof. We prove it by induction. The case $d = 2$ has been proved by Leisen and Lijoi (2011) and it is displayed in Equation (10).

Now, suppose that (13) holds for d with $l \leq d$ distinct values $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_l$ among the $\lambda_1, \ldots, \lambda_d$ with multiplicities $n_1, \ldots, n_l$. For the case of $d+1$, there are two possibilities for λ_{d+1}

a) $\lambda_{d+1} \in \{\lambda_1, \ldots, \lambda_d\}$

b) $\lambda_{d+1} \notin \{\lambda_1, \ldots, \lambda_d\}$

Case a). In the first case, without loss of generality, we can assume that

$\lambda_{d+1} = \lambda_1 = \tilde\lambda_1 \neq \tilde\lambda_l = \lambda_d$. Hence,

$$
\begin{aligned}
&\Phi^{\sigma}_{d+1}(\tilde\lambda_1, ..., \tilde\lambda_l; n_1+1, ..., n_l) \\
&\quad = \frac{1}{\tilde\lambda_l - \tilde\lambda_1}\Big[\Phi^{\sigma+1}_d(\tilde\lambda_1, ..., \tilde\lambda_l; n_1, ..., n_l) - \Phi^{\sigma+1}_d(\tilde\lambda_1, ..., \tilde\lambda_l; n_1+1, n_2, ..., n_{l-1}, n_l-1)\Big] \\
&\quad = \frac{1}{\tilde\lambda_l - \tilde\lambda_1}\left\{\frac{1}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1-1}\cdots\partial^{n_l-1}}{\partial^{n_1-1}\tilde\lambda_1\cdots\partial^{n_l-1}\tilde\lambda_l}\left[\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda})\prod_{i=1}^l \tilde\lambda_i^{n_i-1}\right]\right. \\
&\qquad \left. - \frac{n_l-1}{n_1}\frac{1}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1}\partial^{n_2-1}\cdots\partial^{n_{l-1}-1}\partial^{n_l-2}}{\partial^{n_1}\tilde\lambda_1\partial^{n_2-1}\tilde\lambda_2\cdots\partial^{n_{l-1}-1}\tilde\lambda_{l-1}\partial^{n_l-2}\tilde\lambda_l}\left[\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda})\frac{\tilde\lambda_1}{\tilde\lambda_l}\prod_{i=1}^l \tilde\lambda_i^{n_i-1}\right]\right\} \\
&\quad = \frac{1}{\tilde\lambda_l - \tilde\lambda_1}\left\{\frac{1}{n_1}\frac{1}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1-1}\cdots\partial^{n_l-1}}{\partial^{n_1-1}\tilde\lambda_1\cdots\partial^{n_l-1}\tilde\lambda_l}\left[\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda})\frac{\partial}{\partial\tilde\lambda_1}\left(\tilde\lambda_1^{n_1}\right)\prod_{i=2}^l \tilde\lambda_i^{n_i-1}\right]\right. \\
&\left. -\frac{1}{n_1}\frac{1}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1}\partial^{n_2-1}\cdots\partial^{n_{l-1}-1}\partial^{n_l-2}}{\partial^{n_1}\tilde\lambda_1\partial^{n_2-1}\tilde\lambda_2\cdots\partial^{n_{l-1}-1}\tilde\lambda_{l-1}\partial^{n_l-2}\tilde\lambda_l}\left[\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda})\tilde\lambda_1\prod_{i=1}^{l-1}\tilde\lambda_i^{n_i-1}\frac{\partial}{\partial\tilde\lambda_l}\left(\tilde\lambda_l^{n_l-1}\right)\right]\right\}
\end{aligned}
\tag{14}
$$

Let $\boldsymbol\lambda_{-1} = (\tilde\lambda_2, \dots, \tilde\lambda_l)$. The following identity holds

$$\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda}) = \tilde\lambda_1\phi^{\sigma}_l(\tilde{\boldsymbol\lambda}) + \phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda}_{-1}) \tag{15}$$

Notice that $\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda}_{-1})$ does not depend on $\tilde\lambda_1$. Using this identity, the second term of (14) could be written as

$$
\begin{aligned}
-\frac{1}{\tilde\lambda_l - \tilde\lambda_1}&\left\{\frac{1}{n_1}\frac{1}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde\lambda_1\cdots\partial^{n_l-1}\tilde\lambda_l}\left[\phi^{\sigma+1}_l(\tilde{\boldsymbol\lambda}_{-1})\tilde\lambda_1^{n_1}\prod_{i=2}^l \tilde\lambda_i^{n_i-1}\right]\right. \\
&\left. - \frac{1}{n_1}\frac{\tilde\lambda_l}{\prod_{i=1}^l \Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde\lambda_1\cdots\partial^{n_l-1}\tilde\lambda_l}\left[\phi^{\sigma}_l(\tilde{\boldsymbol\lambda})\tilde\lambda_1\prod_{i=1}^l \tilde\lambda_i^{n_i-1}\right]\right\}
\end{aligned}
\tag{16}
$$

On the other hand, the first term of (14) could be written as

$$\frac{1}{\tilde{\lambda}_l-\tilde{\lambda}_1}\left\{\frac{1}{n_1}\frac{1}{\prod_{i=1}^l\Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l}\left[\phi_l^{\sigma+1}(\tilde{\boldsymbol{\lambda}})\tilde{\lambda}_1^{n_1}\prod_{i=2}^l\tilde{\lambda}_i^{n_i-1}\right]\right.$$
$$\left.-\frac{1}{n_1}\frac{1}{\prod_{i=1}^l\Gamma(n_i)}\frac{\partial^{n_1-1}\cdots\partial^{n_l-1}}{\partial^{n_1-1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l}\left[\tilde{\lambda}_1^{n_1}\prod_{i=2}^l\tilde{\lambda}_i^{n_i-1}\frac{\partial}{\partial\tilde{\lambda}_1}\phi_l^{\sigma+1}(\tilde{\boldsymbol{\lambda}})\right]\right\} \tag{17}$$

By applying the general Leibniz rule and noting that the n_1-th derivative of $\tilde{\lambda}_1^{n_1-1}$ is 0, one gets,

$$\begin{aligned}\frac{\partial^{n_1-1}}{\partial^{n_1-1}\tilde{\lambda}_1}\left[\tilde{\lambda}_1^{n_1}\prod_{i=2}^l\tilde{\lambda}_i^{n_i-1}\frac{\partial}{\partial\tilde{\lambda}_l}\phi_l^{a+1}(\tilde{\boldsymbol{\lambda}})\right] &= \sum_{k=0}^{n_1-1}\binom{n_1-1}{k}\frac{\partial^k}{\partial^k\tilde{\lambda}_1}(\tilde{\lambda}_1^{n_1})\frac{\partial^{n_1-k}}{\partial^{n_1-k}\tilde{\lambda}_l}\phi_l^{a+1}(\tilde{\boldsymbol{\lambda}})\\ &=\tilde{\lambda}_1\sum_{k=0}^{n_1}\binom{n_1}{k}\frac{\partial^k}{\partial^k\tilde{\lambda}_1}(\tilde{\lambda}_l^{n_1-1})\frac{\partial^{n_1-k}}{\partial^{n_1-k}\tilde{\lambda}_1}\phi_l^{a+1}(\tilde{\boldsymbol{\lambda}})\\ &=\tilde{\lambda}_1\frac{\partial^{n_1}}{\partial^{n_1}\tilde{\lambda}_1}\tilde{\lambda}_l^{n_1-1}\phi_l^{a+1}(\tilde{\boldsymbol{\lambda}})\end{aligned}$$

and from identity (15) , equation (17) could be further written as

$$\frac{1}{\tilde{\lambda}_l-\tilde{\lambda}_1}\left\{\frac{1}{n_1}\frac{1}{\prod_{i=1}^l\Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l}\left[\phi_l^{\sigma+1}(\tilde{\boldsymbol{\lambda}}_{-1})\tilde{\lambda}_1^{n_1}\prod_{i=2}^l\tilde{\lambda}_i^{n_i-1}\right]\right.$$
$$\left.-\frac{1}{n_1}\frac{\tilde{\lambda}_1}{\prod_{i=1}^l\Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l}\left[\phi_l^{\sigma}(\tilde{\boldsymbol{\lambda}})\tilde{\lambda}_1\prod_{i=1}^l\tilde{\lambda}_i^{n_i-1}\right]\right\} \tag{18}$$

Combining equation (16) with (18) we get the thesis of case a), i.e.

$$\Phi_{d+1}^a(\tilde{\lambda}_1,...,\tilde{\lambda}_l;n_1+1,...,n_l)=\frac{1}{n_1}\frac{1}{\prod_{i=1}^l\Gamma(n_i)}\frac{\partial^{n_1}\cdots\partial^{n_l-1}}{\partial^{n_1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l}\left[\phi_l^a(\tilde{\boldsymbol{\lambda}})\tilde{\lambda}_1\prod_{i=1}^l\tilde{\lambda}_i^{n_i-1}\right]$$

Case b). Without loss of generality, we assume that $\lambda_{d+1}=\tilde{\lambda}_{l+1}\neq\tilde{\lambda}_l=\lambda_d$.

Hence,

$$\Phi^{\sigma}_{d+1}(\tilde{\lambda}_1, ..., \tilde{\lambda}_l, \tilde{\lambda}_{l+1}; n_1, ..., n_l, 1) =$$
$$\frac{1}{\tilde{\lambda}_l - \tilde{\lambda}_{l+1}}\Big[\Phi^{\sigma+1}_{d}(\tilde{\lambda}_1, ..., \tilde{\lambda}_l; n_1, ..., n_l) - \Phi^{\sigma+1}_{d}(\tilde{\lambda}_1, ..., \tilde{\lambda}_l, \tilde{\lambda}_{l+1}; n_1, ..., n_l - 1, 1)\Big]$$

By working in a similar fashion of "case a)", one gets the thesis for "case b)", i.e.

$$\Phi^{\sigma}_{d+1}(\tilde{\lambda}_1, ..., \tilde{\lambda}_l, \tilde{\lambda}_{l+1}; n_1, ..., n_l, 1) = \frac{1}{\prod_{i=1}^{l}\Gamma(n_i)}\frac{\partial^{n_1-1}\cdots\partial^{n_l-1}}{\partial^{n_1-1}\tilde{\lambda}_1\cdots\partial^{n_l-1}\tilde{\lambda}_l} \times \left[\prod_{i=1}^{l+1}\tilde{\lambda}_i^{n_i-1}\phi^{\sigma}_{l+1}(\tilde{\lambda}_1, ..., \tilde{\lambda}_{l+1})\right]$$

and this concludes the proof. □

4. The Exchangeable Partition Probability Function

In this Section, the exchangeable partition probability function (EPPF) is computed for the vector defined in (8). Let us define the following function

$$g_\rho(q_1, \ldots, q_d; \boldsymbol{\lambda}) = \int_{(0,\infty)^d} y_1^{q_1}\cdots y_d^{q_d} e^{-\psi_{\rho,d}(\boldsymbol{\lambda})}\rho_d(y_1, \ldots, y_d)d\boldsymbol{y} \tag{19}$$

that plays a central role in the determination of the EPPF. We will prove that this function is related with a Lauricella function of fourth kind that is defined as

$$F_D(a, b_1, \ldots, b_d, c, z_1, \ldots, z_d) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_d=0}^{\infty}\frac{(a)_{|\boldsymbol{m}|}(b_1)_{m_1}\cdots(b_d)_{m_d}}{(c)_{|\boldsymbol{m}|}m_1!\cdots m_d!}z_1^{m_1}\cdots z_d^{m_d}$$

for $|z_i| < 1$, $i = 1, \ldots, d$ and $|\boldsymbol{m}| = m_1 + \cdots + m_d$. It is available an integral representation of the Lauricella function of fourth kind, that is

$$F_D(a, b_1, \dots, b_d, c, z_1, \dots, z_d) = \frac{\Gamma(c)}{\Gamma(c - \sum_{i=1}^d b_i) \prod_{i=1}^d \Gamma(b_i)} \times \int_{\Delta_d} (1 - |\boldsymbol{y}|)^{c-1-\sum_{i=1}^d b_i} \prod_{i=1}^d y_i^{b_i - 1} (1 - <\boldsymbol{y}, \boldsymbol{z}>)^{-a} d\boldsymbol{y}$$

Hence, we have the following theorem.

Theorem 1. *Let $g_\rho(q_1, ..., q_d; \boldsymbol{\lambda})$ be defined as (19). Let $I \in \{1, \dots, d\}$ be such that $\lambda_I = \max(\lambda_1, \dots, \lambda_d)$. Hence,*

$$g_\rho(q_1, ..., q_d; \boldsymbol{\lambda}) = \frac{(\sigma)_d}{\Gamma(1-\sigma)} \frac{\Gamma(|\boldsymbol{q}| - \sigma)}{\lambda_I^{|\boldsymbol{q}|-\sigma}} \frac{\prod_{i=1}^d \Gamma(q_i + 1)}{\Gamma(|\boldsymbol{q}| + d)} \times F_D(|\boldsymbol{q}| - \sigma; \boldsymbol{q}_{-I} + \mathbf{1}; |\boldsymbol{q}| + d; \mathbf{1} - \frac{\boldsymbol{\lambda_{-I}}}{\lambda_I})$$

where $\boldsymbol{q}_{-I} + \mathbf{1}$ and $\mathbf{1} - \frac{\boldsymbol{\lambda_{-I}}}{\lambda_I}$ are the vectors of parameters

$$\boldsymbol{q}_{-I} + \mathbf{1} = (q_1 + 1, \dots, q_{I-1} + 1, q_{I+1} + 1, \dots, q_d + 1)$$
$$\mathbf{1} - \frac{\boldsymbol{\lambda_{-I}}}{\lambda_I} = \left(1 - \frac{\lambda_1}{\lambda_I}, ..., 1 - \frac{\lambda_{I-1}}{\lambda_I}, 1 - \frac{\lambda_{I+1}}{\lambda_I}, \cdots, 1 - \frac{\lambda_d}{\lambda_I}\right)$$

Proof. Without loss of generality, suppose that $I = d$. A simple change of variable, namely $z_i = y_i/s$ for $i = 1, ..., d-1$ and $s = |\mathbf{y}|$, yields

$$\begin{aligned} g_\rho(q_1, ..., q_d; \boldsymbol{\lambda}) &= \frac{(\sigma)_d}{\Gamma(1-\sigma)} \int_{\Delta_{d-1}} (1 - |\boldsymbol{z}|)^{q_d} \prod_{i=1}^{d-1} z_i^{q_i} \\ &\quad \times \int_0^{+\infty} s^{|\boldsymbol{q}|-\sigma-1} e^{-s[<\boldsymbol{\lambda}, \boldsymbol{z}> + \lambda_d(1-|\boldsymbol{z}|)]} ds d\boldsymbol{z} \\ &= \frac{(\sigma)_d \Gamma(|\boldsymbol{q}| - \sigma)}{\Gamma(1-\sigma)} \int_{\Delta_{d-1}} \frac{(1 - |\boldsymbol{z}|)^{q_d} \prod_{i=1}^{d-1} z_i^{q_i}}{[<\boldsymbol{\lambda}, \boldsymbol{z}> + \lambda_d(1 - |\boldsymbol{z}|)]^{|\boldsymbol{q}|-\sigma}} d\boldsymbol{z} \end{aligned}$$

and we get the thesis from the integral representation of the Lauricella function of fourth kind. □

As in Leisen and Lijoi (2011), we are considering d groups of data with sample sizes $n_1, \ldots, n_d$ that are partial exchangeable, i.e.

$$\mathbb{P}\left[\boldsymbol{X}^{(1)}_{n_1} \in \times_{i_1=1}^{n_1} A^{(1)}_{i_1}; \ldots; \boldsymbol{X}^{(d)}_{n_d} \in \times_{i_d=1}^{n_d} A^{(d)}_{i_d} \mid (\tilde{p}_1, \ldots, \tilde{p}_d)\right] = \prod_{i_1=1}^{n_1} \tilde{p}_1(A^{(1)}_{i_1}) \times \cdots$$
$$\cdots \times \prod_{i_d=1}^{n_d} \tilde{p}_d(A^{(d)}_{i_d}).$$

with $\boldsymbol{X}^{(i)}_{n_i} = (X^{(i)}_1, \ldots, X^{(i)}_{n_i})$, $i = 1, \ldots, d$. This description of the model implies that the d samples $(X^{(1)}_1, \ldots, X^{(1)}_{n_1}), \ldots, (X^{(d)}_1, \ldots, X^{(d)}_{n_d})$ are independent conditional on $(\tilde{p}_1, \ldots, \tilde{p}_d)$. Given the discrete nature of the random probabilities in (8), there might be ties, i.e. common values among the samples $\boldsymbol{X}^{(i)}_{n_i}$, $i = 1, \ldots, d$. Precisely, let $Z^*_1, \ldots, Z^*_K$ be the distinct values among the $(X^{(1)}_1, \ldots, X^{(1)}_{n_1}), \ldots, (X^{(d)}_1, \ldots, X^{(d)}_{n_d})$. Clearly, $1 \leq K \leq n_1 + \cdots + n_d$. Let $N_{i,j}$ be the number of $X^{(j)}$'s that is equal to Z^*_i, i.e.

$$N_{i,j} = \sum_{h=0}^{n_j} 1_{\{X^{(j)}_h = Z^*_i\}}$$

This means that the data can be described by the set

$$\{K, Z^*_1, \ldots, Z^*_K, (N_{1,1}, \ldots, N_{K,1}), \ldots, (N_{1,d}, \ldots, N_{K,d})\}$$

The *Exchangeable Partition Probability Function* (EPPF) is defined as

$$\Pi^{n_1,\ldots,n_d}_k(\mathbf{n_1}, ..., \mathbf{n_d}) = \mathbb{P}[K = k, \mathbf{N_1} = \mathbf{n_1}, ..., \mathbf{N_d} = \mathbf{n_d}]$$

with $1 \leq k \leq n_1 + \cdots + n_d$ and for vector of non-negative integers such that $\sum_{h=0}^{k} n_{i,h} = n_i$. In the following theorem, an expression of the EPPF is provided and an algorithm for its evaluation will be given in next section.

Theorem 2. *For any positive integers* $n_1, ..., n_d$ *and* k *and vectors* $\mathbf{n_i} = (n_{1,i}, ..., n_{k,i})$ *for* $i = 1, ..., d$ *such that* $\sum_{j=1}^{k} n_{j,i} = n_i$ *and* $n_{j,1} + \cdots + n_{j,d} \geq 1$, *for $i = 1, ..., d$, one has*

$$\Pi_k^{n_1,...,n_d}(\mathbf{n_1}, ..., \mathbf{n_d}) = \frac{\displaystyle\int_{(0,\infty)^d} \lambda_1^{\theta+n_1-1} \cdots \lambda_d^{\theta+n_d-1} e^{-\psi_{\rho,d}(\boldsymbol{\lambda})} \times \prod_{j=1}^{k} g_\rho(n_{j,1}, ..., n_{j,d}; \boldsymbol{\lambda}) d\boldsymbol{\lambda}}{K \displaystyle\prod_{i=1}^{d} (\theta)_{n_i}}$$

Proof. We recall that with $\tilde{\mu}_i$, $i = 1, ..., d$, we denote the i-th σ-stable completely random measure, see Section 2.

$$\tilde{\pi}_k^{n_1,...,n_d}(\mathbf{n_1}, ..., \mathbf{n_d}, d\mathbf{z}) = \frac{\Gamma^d(\theta)}{K \prod_{i=1}^{d} [\tilde{\mu}_i(\mathbb{X})]^{\theta+n_i}} \prod_{j=1}^{k} [\tilde{\mu}_1(dz_j)]^{n_{j,1}} \cdot \cdots \cdot [\tilde{\mu}_d(dz_j)]^{n_{j,d}}$$

for any $k \geq 1$ and $\mathbf{n_i} = (n_{1,i}, ..., n_{k,i})$ such that $\sum_{j=1}^{k} n_{j,i} = n_i$ for $i = 1, ..., d$. We will now show that the probability distribution $\mathbb{E}\,[\tilde{\pi}_k^{n_1,...,n_d}]$ admites a density on $\mathbb{N}^{dk} \times \mathbb{X}^k$ with respect to the product measure $\gamma^{dk} \times \alpha^k$, where γ is the counting measure on the positive integers, and will determine its form. Suppose $C_{\epsilon,x}$ denotes a neighborhood of $x \in \mathbb{X}$ of radius $\epsilon > 0$ and $B_\epsilon = \cup_{j=1}^{k} C_{\epsilon,z_j}$. Then

$$\int_{B_\epsilon} \mathbb{E}\,[\tilde{\pi}_k^{n_1,...,n_d}(\mathbf{n_1}, ..., \mathbf{n_d}, d\mathbf{z})] = \frac{1}{K \prod_{i=1}^{d} (\theta)_{n_i}} \int_{(0,\infty)^d} \lambda_1^{\theta+n_1-1} \cdot \cdots \cdot \lambda_d^{\theta+n_d-1}$$
$$\times \mathbb{E}\left[e^{-\lambda_1 \tilde{\mu}_1(\mathbb{X}) - \cdots - \lambda_d \tilde{\mu}_d(\mathbb{X})} \prod_{j=1}^{k} \prod_{i=1}^{d} \left[\tilde{\mu}_i(C_{\epsilon,z_j})\right]^{n_{j,i}} \right] d\boldsymbol{\lambda}$$

Define $\mathbb{X}_\epsilon$ to be the whole space $\mathbb{X}$ with the neighbourhoods C_{ϵ,z_r} deleted for all $j = 1, ..., k$. By virtue of the independence of the increments of the CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$, the expression above reduces to

$$\frac{1}{K \prod_{i=1}^{d} (\theta)_{n_i}} \int_{(0,\infty)^d} \lambda_1^{\theta+n_1-1} \cdots \lambda_d^{\theta+n_d-1} \mathbb{E}\left[e^{-\lambda_1 \tilde{\mu}_1(\mathbb{X}_\epsilon) - \cdots - \lambda_d \tilde{\mu}_d(\mathbb{X}_\epsilon)} \right] \times \prod_{j=1}^{k} M_{j,\epsilon}(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$$

where, by virtue of Lemma 1 in Appendix,

$$
\begin{aligned}
M_{j,\epsilon}(\boldsymbol{\lambda}) :&= \mathbb{E}\left[e^{-\lambda_1 \tilde{\mu}_1(\mathbb{X}_\epsilon)-\cdots-\lambda_d \tilde{\mu}_d(\mathbb{X}_\epsilon)} \prod_{i=1}^{d} \left[\tilde{\mu}_i(C_{\epsilon,z_j})\right]^{n_{j,i}}\right] \\
&= \alpha(C_{\epsilon,z_j}) e^{-\alpha(C_{\epsilon,z_j})\psi_{\rho,d}(\boldsymbol{\lambda})} g_\rho(n_{j,1},...,n_{j,d};\boldsymbol{\lambda}) + o(\alpha(C_{\epsilon,z_j}))
\end{aligned}
$$

This shows that $\mathbb{E}[\tilde{\pi}^k]$ admits a density with respect to $\gamma^{dk} \times \alpha^k$ and it is given by

$$
\frac{1}{K\prod_{i=1}^{d}(\theta)_{n_i}} \int_{(0,\infty)^d} \lambda_1^{\theta+n_1-1} \cdots \lambda_d^{\theta+n_d-1} e^{-\psi_{\rho,d}(\boldsymbol{\lambda})} \times \prod_{j=1}^{k} g_\rho(n_{j,1},...,n_{j,d};\boldsymbol{\lambda}) d\boldsymbol{\lambda}
$$

□

5. Numerical illustrations

In this section we provide an automatic Markov Chain Monte Carlo algorithm (MCMC) for evaluating the EPPF when $\theta > 1$. In particular, in the first part we will focus on the bidimensional case for comparing the performance of the algorithm vs the matlab integrator. Indeed, as shown in Leisen and Lijoi (2011), in the bivariate case, the EPPF reduces to a one dimensional integral in the interval $[0,1]$ and, in this case, the matlab integrator is very accurate. When the dimension d of the vector is greater than 2, it is not possible to reduce the dimension of the integral and another computational approach must be used. In the MCMC setting it is possible to compute integrals of this type:

$$
I = \frac{\int F(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x}}{Z} \tag{20}
$$

where

$$
Z = \int \pi(\boldsymbol{x})d\boldsymbol{x}
$$

is the normalizing constant of the unnormalized *target* distribution π and F a objective funtion. This means that

$$
\tilde{\pi}(x) = \frac{\pi(x)}{Z}
$$

is a probability distribution. The Metropolis-Hastings algorithm is able to construct a Markov Chain $(X_i)_{i\geq 0}$ that has $\tilde{\pi}$ as stationary distribution and, an estimator of the integral (20) is

$$I \cong \frac{1}{N}\sum_{i=1}^{N} F(X_i)$$

Since the constant K in the EPPF acts naturally as a normalizing constant, in our case, the unnormalized target is

$$\pi(\boldsymbol{\lambda}) = e^{-\psi_{\rho,d}(\boldsymbol{\lambda})}\prod_{i=1}^{d}\lambda_i^{\theta-1} \tag{21}$$

and the function F is

$$F(\boldsymbol{\lambda}) = \left(\prod_{i=1}^{d}\frac{\lambda_i^{n_i}}{(\theta)_{n_i}}\right)\left(\prod_{j=1}^{k} g_\rho(n_{j,1},...,n_{j,d};\boldsymbol{\lambda})\right)$$

For a non-expert reader, the general steps (in our case) of the Metropolis-Hastings, are displayed in Algorithm 1.

Algorithm 1. *Random Walk Metropolis-Hastings for the EPPF*

Suppose that $\boldsymbol{\lambda}^{(t)} = (\lambda_1^{(t)}, \ldots, \lambda_d^{(t)})$

1. *Draw* $\boldsymbol{\lambda}' = (\lambda_1', \ldots, \lambda_d')$ *from a proposal distribution* $Q(\cdot|\boldsymbol{\lambda}^{(t)})$.
2. *Set* $\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}'$ *with probability*

$$\alpha(\boldsymbol{\lambda}^{(t)}, \boldsymbol{\lambda}') = \min\left\{1, \frac{\pi(\boldsymbol{\lambda}')Q(\boldsymbol{\lambda}^{(t)}|\boldsymbol{\lambda}')}{\pi(\boldsymbol{\lambda}^{(t)})Q(\boldsymbol{\lambda}'|\boldsymbol{\lambda}^{(t)})}\right\}$$

and $\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)}$ *with probability* $1 - \alpha(\boldsymbol{\lambda}^{(t)}, \boldsymbol{\lambda}')$

We set the proposal distribution as

$$Q(\cdot|\boldsymbol{\lambda}^{(t)}) = q_1(\cdot|\lambda_1^{(t)}) \cdots q_d(\cdot|\lambda_d^{(t)})$$

Since θ regulates the shape of the proposal distribution, we need to choose $q_i(\cdot|\lambda_i^{(t)})$ according to its value.

If $\theta \leq 1$ then $q_i(\cdot|\lambda_i^{(t)})$, $i = 1, \ldots, d$ is a Weibull Distribution, otherwise if $\theta > 1$, then $q_i(\cdot|\lambda_i^{(t)})$, $i = 1, \ldots, d$, is the density of a truncated Normal distribution in the interval $[0, \infty)$ with mean $\lambda_i^{(t)}$ and standard deviation σ_i. When $\theta > 1$, a guideline for setting the standard deviations $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ could be

$$\boldsymbol{\sigma} = \text{Argmax } \pi(\boldsymbol{\lambda})$$

On the other hand, when $\theta \leq 1$, a guideline could be to set the shape parameter of the Weibull distribution equal to σ. Since the heaviness of the tails increases as σ decreases, we suggest the following empirical rule for the scale parameter: if $\sigma \geq 0.5$ then set it equal to 1 otherwise, set it at least 10.

In the set of experiments, we run 20 chains for 20000 iterations with a burn in period of 5000 iterations.

First of all, we focus on the bidimensional case and we test the performance of our MCMC algorithm. Note that the expression of the EPPF in Theorem 2 is done by 2 integrals, one at the numerator and the constant K at the denominator. When $d = 2$, in a similar fashion of Leisen and Lijoi (2011), both integrals can be reduced to one dimensional integrals in the interval $[0, 1]$ and evaluated accurately through the standard one-dimensional matlab integrator. This allows a comparison with our MCMC algorithm and the results are displayed in Table 1 for different values of θ and σ.

As we can see in Table 1, our algorithm has a good level of accuracy compared with the matlab integrator. In some cases, it performs better, i.e when $(\theta = 5, \sigma = 0.25)$ and $(\theta = 5, \sigma = 0.8)$. This inaccuracy of the matlab integrator is due to the "explosion" of the values of both integrals at the numerator and denominator. Our algorithm, is unaffected by this problem and also by dimensional problems. In a slightly more advanced case, the same comparison is done when $n_1 = 2$ and $n_2 = 1$, see Table 2. Although these examples are very simple, we can do some comments about the clustering behaviour of such a vector. Precisely, if θ or σ increases then the number of clusters increases. Moreover, in table 2 it can be observed that when $\theta = 0.5$, the probability of the configuration with k=1 cluster changes its trend, as σ

		Numerical Integrator		MCMC	
θ	σ	$k = 1$	$k = 2$	$k = 1$	$k = 2$
0.5	0.25	0.294	0.706	0.2985 ± 0.0033	0.7075 ± 0.2022
0.5	0.5	0.1894	0.8106	0.1897 ± 0.0013	0.7947 ± 0.0158
0.5	0.8	0.0729	0.927	0.0729 ± 0.0005	0.9150 ± 0.0142
1	0.25	0.1862	0.8138	0.1867 ± 0.0034	0.8096 ± 0.0036
1	0.5	0.1215	0.8785	0.1218 ± 0.0008	0.8818 ± 0.0091
1	0.8	0.0475	0.9525	0.0476 ± 0.0003	0.9521 ± 0.0130
3	0.25	0.0734	0.9053	0.0759 ± 0.0048	0.9154 ± 0.0882
3	0.5	0.0496	0.9504	0.0498 ± 0.0012	0.9397 ± 0.0486
3	0.8	0.0197	0.981	0.0198 ± 0.0003	0.9769 ± 0.050
5	0.25	**0.0423**	**0.8585**	0.0472 ± 0.0004	0.9593 ± 0.0327
5	0.5	0.0303	0.9439	0.0313 ± 0.0003	0.9577 ± 0.0252
5	0.8	**0.0126**	**1.0061**	0.0124 ± 0.0001	0.9748 ± 0.0201

Table 1: Matlab integrator vs MCMC assuming $n_1 = 1$ and $n_2 = 1$

increases, compared with the first 2 configurations with k=2 clusters. Indeed, when $\sigma = 0.25$, the first one has higher probability than the second one and, conversely, when $\sigma = 0.8$, the first one has lower probability than the second one. This suggests that an higher σ encourages an higher number of clusters and a lower sigma discourages the creation of new clusters.

Anyway, for a better understanding of the qualitative behaviour, we need to test the prior on more sophisticated configurations.

We consider the three dimensional case where $n_1 = 40, n_2 = 20$ and $n_3 = 30$. In Figure 1 is displayed, for some values of θ and σ, the log scale EPPF for two different clustering behaviours. On the left hand side is plotted the k=1 cluster case and the right hand side, is plotted a configuration with k=3 clusters and multiplicities $\mathbf{n_1} = (10, 0, 0), \mathbf{n_2} = (10, 10, 10)$ and $\mathbf{n_3} = (20, 10, 20)$.

The probability of k=1 cluster decreases as σ increases and also together with the increase of θ, which evidently suggests that lower σ and lower θ depresses the creation of new clusters. In addition, the comparison of the probabilities of the two cases implies that the first case, which has only 1 cluster, happens with a enormously bigger chance than the second one.

		Numerical Integrator			MCMC		
θ	σ	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
			0.1225			0.1227 ± 0.0069	
0.5	0.25	0.1715	0.1225	0.353	0.1700 ± 0.0043	0.1227 ± 0.0069	0.3503 ± 0.0295
			0.2295			0.2285 ± 0.0124	
			0.0947			0.0959 ± 0.004	
0.5	0.5	0.0947	0.0947	0.5404	0.095 ± 0.0016	0.0959 ± 0.004	0.5478 ± 0.0764
			0.1755			0.1753 ± 0.0117	
			0.0438			0.0439 ± 0.0013	
0.5	0.8	0.0292	0.0438	0.8034	0.0289 ± 0.0004	0.0439 ± 0.0013	0.8037 ± 0.0341
			0.0789			0.0795 ± 0.0023	
			0.0574			0.0570 ± 0.0033	
3	0.25	0.0161	0.0574	0.7355	0.0162 ± 0.0004	0.0570 ± 0.0033	0.7331 ± 0.0798
			0.1124			0.1136 ± 0.0072	
			0.0403			0.0398 ± 0.0013	
3	0.5	0.0093	0.0403	0.8316	0.0092 ± 0.0001	0.0398 ± 0.0013	0.8260 ± 0.07
			0.0785			0.0769 ± 0.0025	
			0.0167			0.0164 ± 0.001	
3	0.8	0.0015	0.0167	0.9319	0.0029 ± 0.00001	0.0164 ± 0.001	0.9154 ± 0.0787
			0.0282			0.0314 ± 0.0007	

Table 2: Matlab integrator vs MCMC assuming $n_1 = 2$ and $n_2 = 1$

Log EPPF in three dimensions

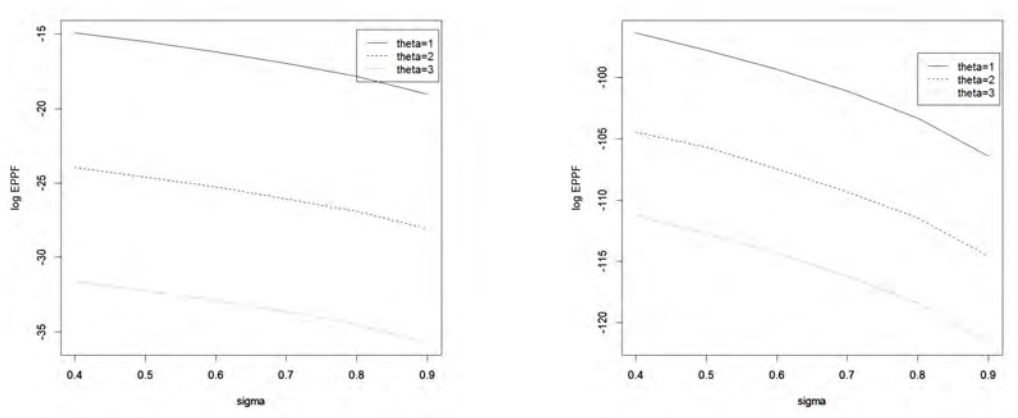

Figure 1: Left column: the log EPPF of k=1 cluster. Right column: the log EPPF of designated k=3 clusters

We conclude the section by pointing out to the reader that similar trends has been observed on higher dimensional cases.

6. Conclusions

In this paper, we extended the bivariate vector of Leisen and Lijoi (2011) to the multivariate setting. We derived some interesting theorical results, especially compared with previous literature. In addition, a new MCMC algorithm has been introduced for evaluating the EPPF since its expression is not explicit. The performance of the algorithm are tested as well as the clustering behavior of such a prior.

Appendix

Lemma 1. *Let* $(\tilde{\mu}_1, ..., \tilde{\mu}_d)$ *be a vector of CRMs with Laplace exponent* $\psi_{\rho,d}(\boldsymbol{\lambda})$. *If* $C_\epsilon \in \mathscr{X}$ *is such that diam*(C_ϵ)$\downarrow 0$ *as* $\epsilon \downarrow 0$, *then*

$$\mathbb{E}\left[e^{-\lambda_1\tilde{\mu}_1(C_\epsilon)-\cdots-\lambda_d\tilde{\mu}_d(C_\epsilon)}\prod_{i=1}^{d}\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right] =$$
$$(-1)^{q_1+\ldots+q_d-1}\alpha(C_\epsilon)e^{-\alpha(C_\epsilon)\psi_{\rho,d}(\boldsymbol{\lambda})} \times \frac{\partial^{q_1+\ldots+q_d}}{\partial\lambda_1^{q_1}\cdots\partial\lambda_d^{q_d}}\psi_{\rho,d}(\boldsymbol{\lambda}) + o(\alpha(C_\epsilon))$$

as $\epsilon \downarrow 0$.

Proof. The proof follows from a simple application of a multivariate version of the Faá di Bruno formula.

$$\mathbb{E}\left[e^{-\lambda_1\tilde{\mu}_1(C_\epsilon)-\cdots-\lambda_d\tilde{\mu}_d(C_\epsilon)}\prod_{i=1}^{d}\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right] = (-1)^{q_1+\ldots+q_d}\frac{\partial^{q_1+\ldots+q_d}}{\partial\lambda_1^{q_1}\cdots\partial\lambda_d^{q_d}}e^{-\alpha(C_\epsilon)\psi_{\rho,d}(\boldsymbol{\lambda})}$$

The right-hand side above coincides with

$$e^{-\alpha(C_\epsilon)\psi_{\rho,d}(\boldsymbol{\lambda})}q_1!\cdots q_d!\sum_{k=1}^{q_1+\ldots+q_d}(-1)^k[\alpha(C_\epsilon)]^k\times$$
$$\sum_{j=1}^{q_1+\ldots+q_d}\sum_{p_j(q_1,\ldots,q_d,k)}\prod_{i=1}^{j}\frac{1}{\beta_i!(s_{1,i}!\cdots s_{d,i}!)^{\beta_i}}\left(\frac{\partial^{s_{1,i}+\ldots+s_{d,i}}}{\partial\lambda_1^{s_{1,i}}\cdots\partial\lambda_d^{s_{d,i}}}\psi_{\rho,d}(\boldsymbol{\lambda})\right)^{\beta_i}$$

where $p_j(q_1,...,q_d,k)$ is the set of vectors $(\boldsymbol{\beta}, \boldsymbol{s_1}, ..., \boldsymbol{s_j})$ with $\boldsymbol{\beta}= (\beta_1,...,\beta_j)$ a vector whose positive coordinates are such that $\sum_{i=1}^{j}\beta_i = k$ and the $\boldsymbol{s_i} = (s_{1,i},...,s_{d,i})$ are vectors such that $\mathbf{0} \prec \mathbf{s_1} \prec \cdots \prec \mathbf{s_j}$. Obviously, in the previous sum, all terms with $k \geq 2$ are $o(\alpha(C_\epsilon))$ as $\epsilon \downarrow 0$. □

Furthermore, if we suppose that the Lévy measure is of finite variation, i.e.
$\int_{\|\mathbf{y}\|\leq 1}\|\mathbf{y}\|\rho_d(y_1,...,y_d)dy_1\cdots dy_d < \infty$ where $\|\mathbf{y}\|$ stands for the Euclidean norm of the vector $\mathbf{y} = (y_1,...,y_d)$, then one also has $\int_{\|\mathbf{y}\|\leq 1}y_1^{n_1}\cdots y_d^{n_d}\rho_d(y_1,...,y_d)dy_1\cdots dy_d < \infty$ for any $n_i,\ i = 1,...,d$ positive integers.

$$\mathbb{E}\left[e^{-\lambda_1\tilde{\mu}_1(C_\epsilon)-\cdots-\lambda_d\tilde{\mu}_d(C_\epsilon)}\prod_{i=1}^{d}\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right] = \alpha(C_\epsilon)e^{-\alpha(C_\epsilon)\psi_{\rho,d}(\boldsymbol{\lambda})}g_\rho(q_1,...,q_d;\boldsymbol{\lambda}) + o(\alpha(C_\epsilon))$$

as $\epsilon \downarrow 0$, for any $\lambda_i > 0$, $i = 1,...,d$, where

$$g_\rho(q_1,...,q_d;\boldsymbol{\lambda}) = \int_{(0,\infty)^d}y_1^{q_1}\cdots y_d^{q_d}e^{-\lambda_1y_1-\cdots-\lambda_dy_d}\rho_d(y_1,...,y_d)dy_1\cdots dy_d$$

References

G.M. Constantines, T.H. Savits (1996). A multivariate version of the Faa di Bruno formula. *Trans. Amer. Math. Soc.* **348**, 503–520.

R. Cont, P. Tankov (2004). *Financial modelling with jump processes.* Chapman & Hall/CRC, Boca Raton, FL.

D.J. Daley, D. Vere-Jones (2003). *An introduction to the theory of point processes. Vol. 1.* Springer, New York.

M. De Iorio, P. Müller, G.L. Rosner, S.N. MacEachern (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99**, 205–215.

Dunson, D.B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian nonparametrics* (Holmes, C.C., Hjort, N.L., Müller, P. and Walker, S.G., Eds.), 223–273, Cambridge University Press, Cambridge.

I. Epifani, A. Lijoi (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica* **20**, 1455–1484.

T.S. Ferguson (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

I. S. Gradshteyn, J.M. Ryzhik (2007). *Table of Integrals, Series, and Products.* 7th Ed. Academic Press, New York.

J. E. Griffin and M. F. J. Steel (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association* **101**, 179–194.

S. J. Hatjispyrosa, T. N. Nicoleris, and S. G. Walker (2011). Dependent mixtures of Dirichlet processes. *Computational Statistics and Data Analysis*, **55**, 2011–2025.

H. Ishwaran, L.F. James (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.* **96**, 161–173.

H. Ishwaran, M. Zarepour (2009). Series representations for multivariate generalized gamma processes via a scale invariance principle. *Statist. Sinica* **19**, 1665–1682.

J. Kallsen, P. Tankov (2006). Characterization of dependence of multidimensional Lévy processes using Lèvy copulas. *J. Multivariate Anal.* **97**, 1551–1572.

J.F.C. Kingman (1975). Kingman, J. F. C. (1975). Random discrete distributions (with discussion). *J. Roy. Statist. Soc. Ser. B* **37**, 1–22.

J.F.C. Kingman (1993). *Poisson processes.* Oxford University Press, Oxford.

M. Kolossiatis, J. Griffin, and M. F. J. Steel (2013). On Bayesian nonparametric modelling of two correlated distributions. *Statistics and Computing*, **23**, 1–15.

F. Leisen, A. Lijoi (2011). Vectors of Poisson-Dirichlet processes. *J. Multivariate Anal.* **102**, 482–495.

F. Leisen, A. Lijoi and D. Spano (2013). A Vector of Dirichlet processes. *Electronic Journal of Statistics* **7**, 62–90.

A. Lijoi, R.H. Mena, I. Prünster (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69**, 715–740.

Lijoi, A., Nipoti, B. and Prünster (2011). Bayesian inference with dependent normalized completely random measures. *Carlo Alberto Notebboks* **224**

A. Lijoi, I. Prünster (2010). Beyond the Dirichlet process. *Bayesian Nonparametrics* (Holmes, C.C., Hjort, N.L., Müller, P. and Walker, S.G., Eds.), 80–136, Cambridge University Press, Cambridge.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12**, 351–357.

S.N. MacEachern (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

J. Pitman (1995). Exchangeable and partially exchangeable random partitions *Probab. Theory Related Fields.* **102**, 145–158.

J. Pitman (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T.S. Ferguson, L.S. Shapley and J.B. MacQueen, Eds.). IMS Lecture Notes Monogr. Ser., Vol. **30**. Inst. Math. Statist., Hayward, 245–267.

J. Pitman (2006). *Combinatorial Stochastic Processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII 2002. Lecture Notes in Mathematics **1875**. Springer, Berlin.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.

E. Regazzini, A. Lijoi, I. Prünster (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31**, 560–585.

E. Regazzini, V.V. Sazonov (2000). Approximation of distributions of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian statistical inferences. *Theory Probab. Appl.* **45**, 93–110.

D. Spanò, A. Lijoi (2012). Canonical correlations for dependent gamma processes. *CRiSM Working Paper* **12–12**, University of Warwick, U.K.

Teh, Y.W. and Jordan, M.I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian nonparametrics* (Holmes, C.C., Hjort, N.L., Müller, P. and Walker, S.G., Eds.), 158–207, Cambridge University Press, Cambridge.